Week 10b: Activity Key

Last Time

- Fitting GP models
- Kriging with plug-in estimates

This week

- Stationarity
- Locations close are more similar, but how? -> Variograms

Stationarity

Assume the spatial process, $Y(s)$ has a mean, $\mu(s)$, and that the variance of $Y(s)$ exists everywhere.

The process is **strictly stationary** if: for any $n \geq 1$, any set of n sites $\{s_1, ..., s_n\}$ and any $h \in \mathcal{R}^r$ (typically r =2), the distribution of $(Y(s_1). ..., Y(s_n))$ is the same as $(Y(s_1 + h) ..., Y(s_n + h)$

Weak stationarity, or second-order stationarity of a spatial process, requires a constant mean and covariance that is a function of h, $Cov(Y(s), Y(s + h)) = C(h)$, where $C(h)$ is a covariance function that only requires the separation vector h.

Typically with spatial models, the process is assumed to be mean zero as covariates explain the mean structure. So second-order stationarity is primarily focused on the covariance structure.

A third kind of stationarity, known as intrinsic stationarity, describes the behavior of differences in the spatial process, rather than the data.

Intrinsic stationarity assumes $E[Y(s+h) - Y(s)] = 0$ then define

$$E[Y(s+h) - Y(s)]^2 = Var(Y(s+h) - Y(s)) = 2\gamma(h),$$

which only works, and satisfies intrinsic stationarity, if the equation only depends on h.

Variograms

Variograms, defined as $2\gamma(h)$ are often used to visualize spatial patterns:

- If the distance between points is small, the variogram is expected to be small

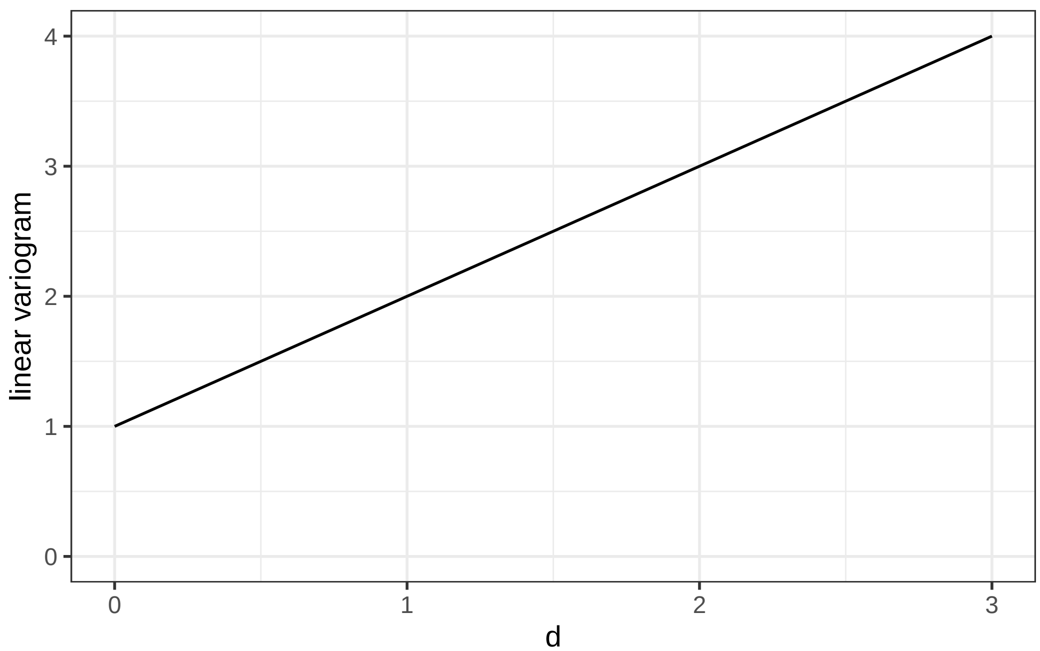
- As the distance between points increases, the variogram increases

There is a mathematical link between the covariance function $C(h)$ and the variogram $2\gamma(h)$.

Linear semivariogram

$$\gamma(d) = \begin{cases} \tau^2 + \sigma^2 d \text{ if } d > 0 \\ 0 \text{ otherwise} \end{cases}$$

```
tau.sq <- 1
sigma.sq <- 1
d <- seq(0,3, by =.01)
lin.gam <- tau.sq + sigma.sq * d
lin.var <- data.frame(d=d, lin.gam = lin.gam )
ggplot(data = lin.var, aes(x=d, y=lin.gam)) +
  geom_line() +
  ylim(0,4) +
  ylab('linear variogram') +
  theme_bw()
```



Nugget, sill, partial-sill, and range

Nugget is defined as $\gamma(d): \; d \rightarrow 0_+$

Sill is defined as $\gamma(d): \; d \rightarrow \infty$

Partial sill is defined as the sill $-$ nugget

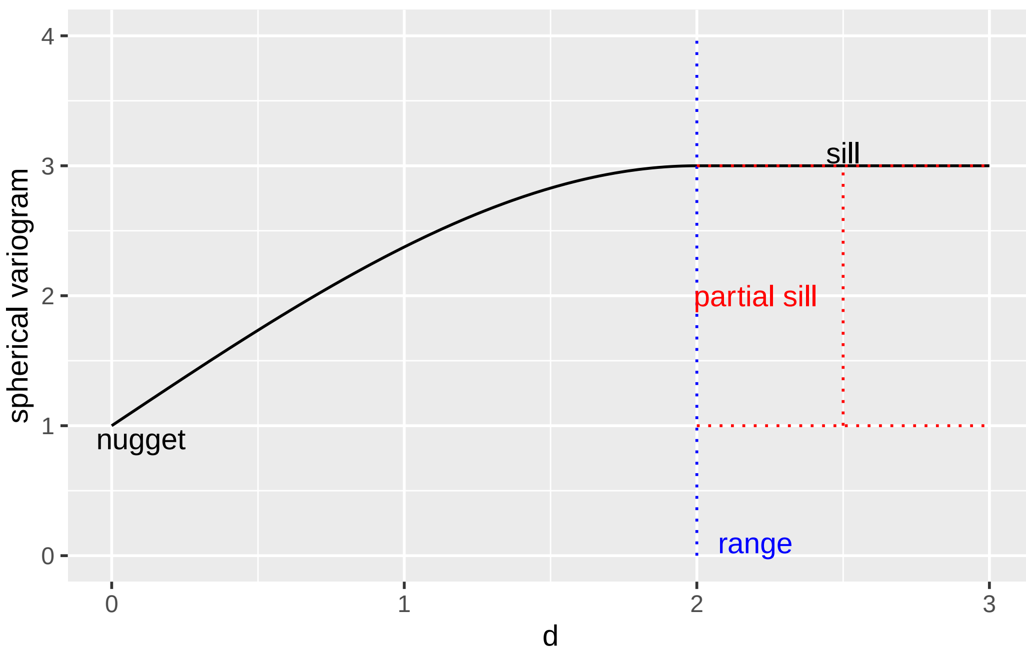
Range is defined as the first point where $\gamma(d)$ reaches the sill. Range is sometimes $(1/\phi)$ and sometimes ϕ depending on the paremeterization.

Spherical semivariogram:

$$\gamma(d) = \begin{cases} \tau^2 + \sigma^2 \text{ if } d \geq 1/\phi \\ \tau^2 + \sigma^2 \left[\frac{3\phi d}{2} - \frac{1}{2}(\phi d)^3\right] \\ 0 \text{ otherwise} \end{cases}$$

- Sketch, or generate in R, a spherical semivariogram
- On this figure label the nugget, sill, partial sill, and range.

Spherical semivariogram: Solution



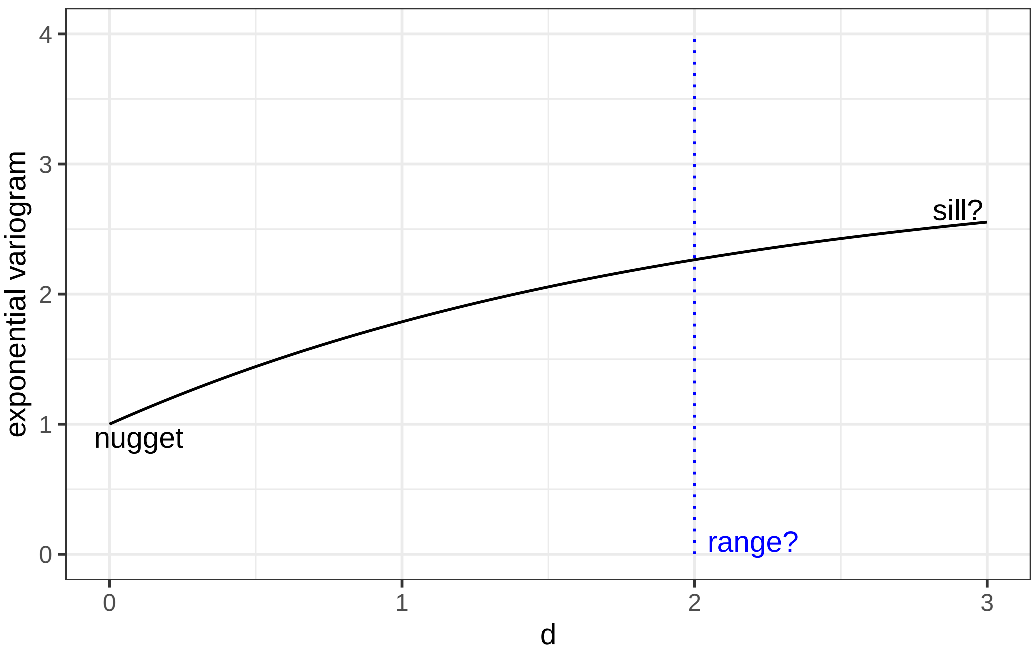
Exponential

We saw the exponential covariance earlier in class, what is the mathematical form of the covariance?

$$C(d) = \begin{cases} \tau^2 + \sigma^2 \text{ if } d = 0 \\ \sigma^2 \exp(-d/\phi) \text{ if } d > 0 \end{cases}$$

The variogram is

$$\gamma(d) = \begin{cases} \tau^2 + \sigma^2(1 - \exp(-d/\phi)) \text{ if } d > 0 \\ 0 \text{ otherwise} \end{cases}$$



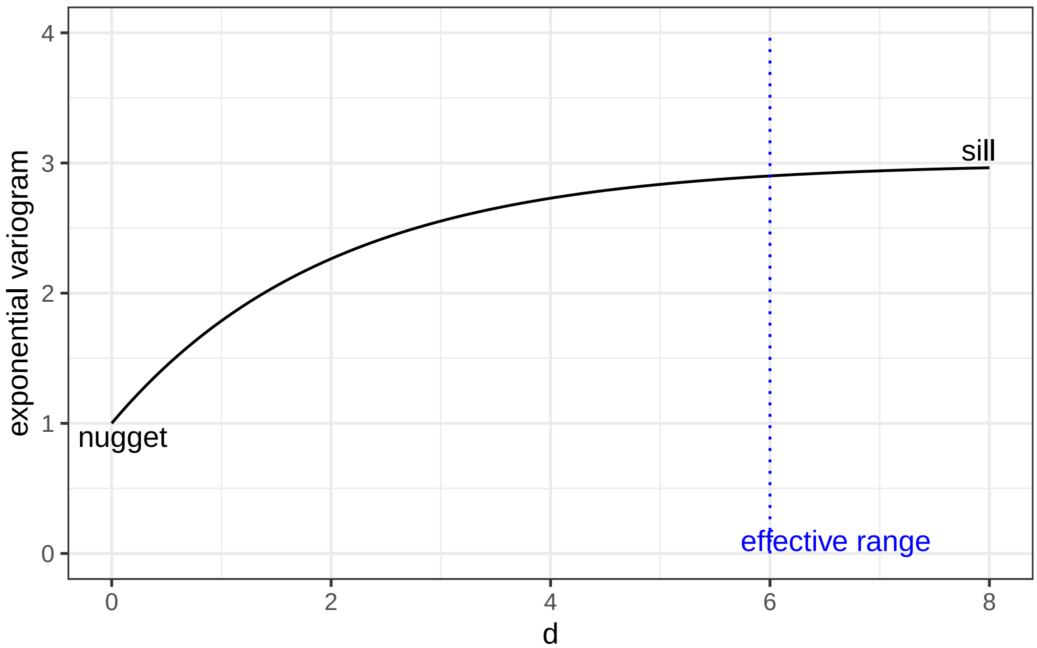
The sill is only reached asymptotically when

$$\lim_{d \to \infty} exp(-d/\phi) \to 0$$

Thus, the range, defined as distance where semivariogram reaches the sill is ∞

The *effective range* is defined as the distance where there is *effectively* no spatial structure. Generally this is determined when the correlation is .05.

$$\begin{aligned} \exp(-d_o/\phi) &= .05 \\ d_0/\phi &= -\log(.05) \\ d_0 &\approx 3\phi \end{aligned}$$



More Semivariograms: Equations

Gaussian:

$$\gamma(d) = \begin{cases} \tau^2 + \sigma^2(1 - \exp(-\phi^2 d^2)) \text{ if } d > 0 \\ 0 \;\; \text{otherwise} \end{cases}$$

Powered Exponential:

$$\gamma(d) = \begin{cases} \tau^2 + \sigma^2(1 - \exp(-|\phi d|^p)) \text{ if } d > 0 \\ 0 \;\; \text{otherwise} \end{cases}$$

Matérn:

$$\gamma(d) = \begin{cases} \tau^2 + \sigma^2 \left[1 - \frac{(2\sqrt{\nu} d\phi)^\nu}{2^{\nu-1}\Gamma(\nu)} K_\nu(2\sqrt{\nu} d\phi)\right] \;\; \text{if } d > 0 \\ 0 \;\; \text{otherwise} \end{cases} ,$$

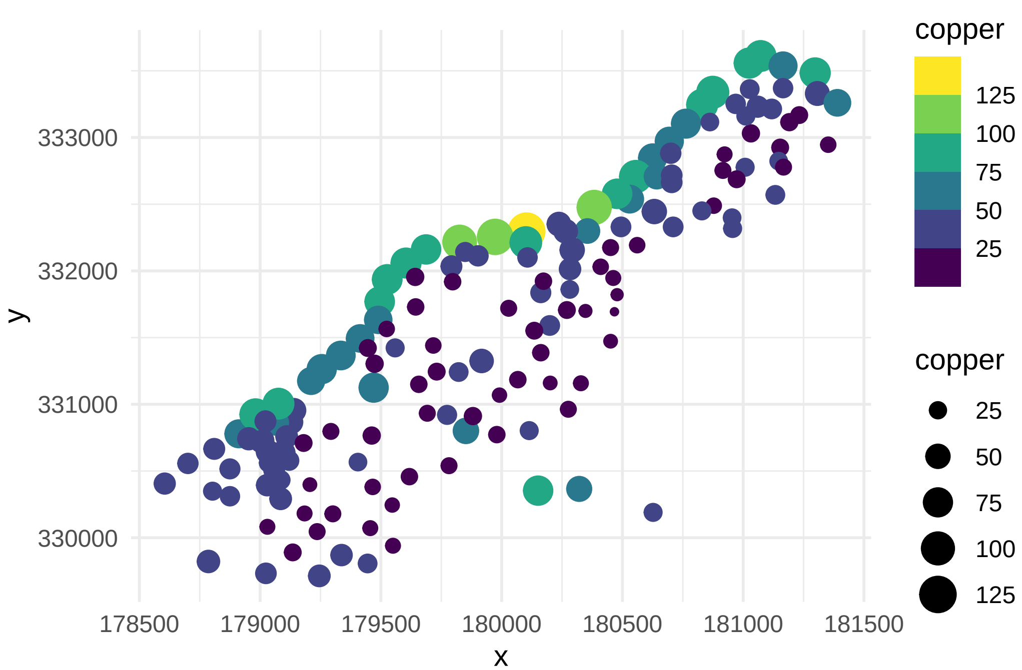
where K_ν is a modified Bessel function and $\Gamma()$ is a Gamma function.

#Variogram Creation: How?

```r
data(meuse)
meuse.small <- meuse %>% select(x, y, copper) %>% as_tibble()
meuse.small %>% head(15) %>% kable()
```

x	y	copper
181072	333611	85
181025	333558	81
181165	333537	68
181298	333484	81
181307	333330	48
181390	333260	61
181165	333370	31
181027	333363	29
181060	333231	37
181232	333168	24
181191	333115	25
181032	333031	25
180874	333339	93
180969	333252	31
181011	333161	27

```
meuse |>
  ggplot(aes(x = x , y=y, size = copper, color = copper)) +
  geom_point() +
  scale_color_viridis_b() +
  theme_minimal()
```



Variogram Creation: Steps

1. Calculate distances between sampling locations
2. Choose grid for distance calculations
3. Calculate empirical semivariogram

$$\hat{\gamma}(d_k) = \frac{1}{2N(d_k)} \sum_{s_i, s_i, \in N(d_k)} \left[Y(s_i) - Y(s_j)\right]^2,$$

   where

$$N(d_k) = \{(s_i, s_j) : ||s_i - s_j|| \in I_k\}$$

   and I_k is the k^{th} interval.
4. Plot the semivariogram

Variogram Creation: How - Step 1

Calculate Distances between sampling locations

```
#dist(meuse.small)
dist.mat <- dist(meuse.small %>% select(x,y))
```

Variogram Creation: How - Step 2

Choose grid for distance calculations

```
cutoff <- max(dist.mat) / 3 # default maximum distance
num.bins <-  15
bin.width <- cutoff / 15
```

Variogram Creation: How - Step 3

Calculate empirical semivariogram

```
dist.sq <- dist(meuse.small$copper)^2

vario.dat <- data.frame(dist = as.numeric(dist.mat),
                        diff = as.numeric(dist.sq)) %>%
  mutate(bin = floor(dist / bin.width) + 1) %>%
  filter(bin < 16) %>%
  group_by(bin) %>%
```

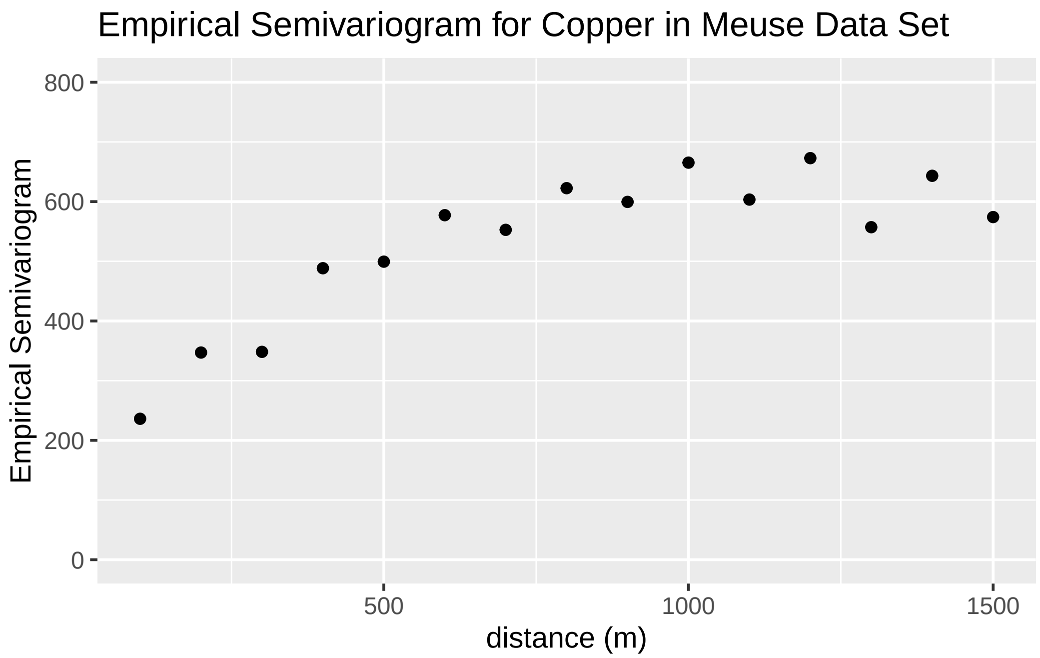
```
  summarize(emp.sv = .5 * mean(diff))
vario.dat
```

```
# A tibble: 15 x 2
     bin emp.sv
   <dbl>  <dbl>
 1     1   236.
 2     2   347.
 3     3   348.
 4     4   488.
 5     5   499.
 6     6   577.
 7     7   553.
 8     8   623.
 9     9   600.
10    10   665.
11    11   603.
12    12   673.
13    13   557.
14    14   643.
15    15   574.
```

Variogram Creation: How - Step 4

Plot empirical semivariogram

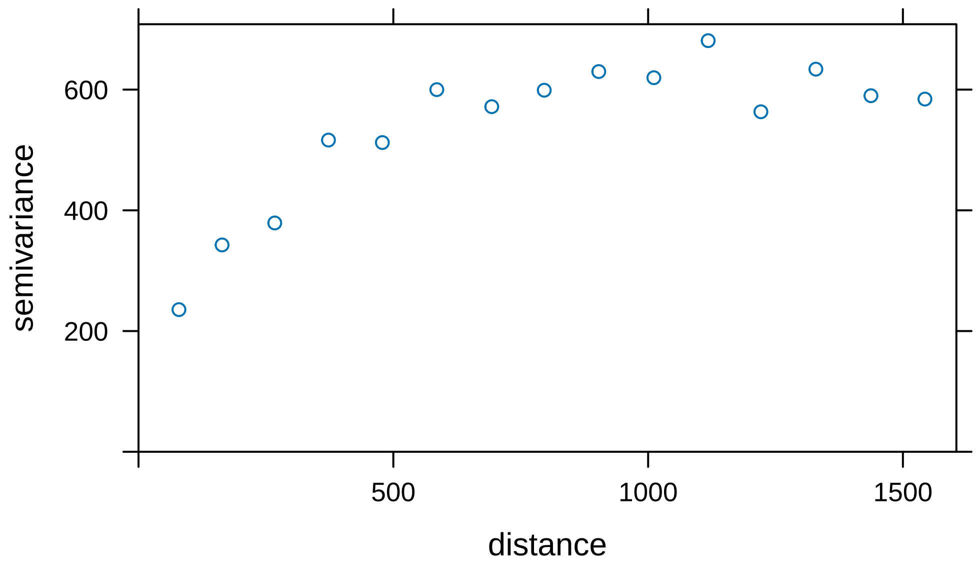
```
vario.dat %>%
  ggplot(aes(x=bin, y= emp.sv)) +
  geom_point() + xlim(0,15) + ylim(0,800) +
  scale_x_continuous(breaks=c(5, 10, 15), labels =c('500','1000','1500')) +
  ggtitle('Empirical Semivariogram for Copper in Meuse Data Set') +
  xlab('distance (m)') + ylab('Empirical Semivariogram')
```



Now given this empirical semivariogram, how to we choose a semivariogram (and associated covariance structure) and estimate the parameters in that function??

Variogram Creation: R function

```
coordinates(meuse) = ~x+y
variogram(copper~1, meuse) %>% plot()
```


600
400
200
semivariance
500
1000
1500
distance